

S&TR January 2021 Antimicrobial Algorithm

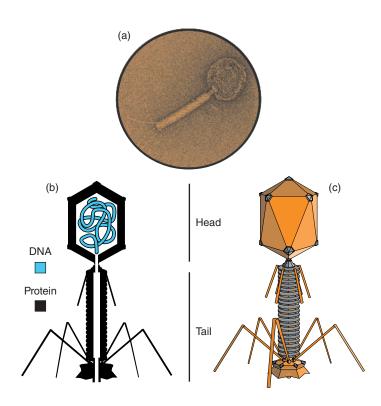
N 2015 while vacationing in Egypt, Tom Patterson contracted a drug-resistant strain of *Acinetobacter baumannii*, a bacterium known to infect as many as 8,500 and kill 700 people in the United States every year. After a battery of antibiotic treatments, recurring bouts of septic shock, and slipping in and out of a coma over the course of several months in what his doctors called, "the worst infection on the planet," Patterson was transferred to the University of California, San Diego (UCSD) where clinicians are familiar with Acinetobacter infections brought home by U.S. forces returning from the Middle East. Running out of options and seeing her husband's condition continue to deteriorate, Patterson's wife, Steffanie Strathdee, an infectious disease epidemiologist at UCSD, proposed an unconventional treatment: bacteriophage therapy. Discovered in the early 1900s, bacteriophage therapy was an emergent, yet controversial means of treating bacterial infections before penicillin took the world by storm. Patterson's treatment would involve an experimental cocktail of microbial viruses to attack the bacteria infecting his body. Within a few days of intravenous administration, Patterson awoke from his coma and began to recover.

Patterson was the first person in North America to receive intravenous bacteriophage therapy to treat a systemic bacterial infection, and his dramatic recovery from what appeared to be an inevitable death sentence helped spark a renewed interest in a potential treatment, which had largely been eschewed by the medical community in the United States and Europe.

To successfully combat increasing antibiotic resistance and treat challenging bacterial infections like Patterson's, scientists in the Forensic Science Center (FSC) at Lawrence Livermore have partnered with San Diego State University and UCSD to advance bacteriophage therapy. Using Livermore's high-performance computing resources and a novel computational algorithm called "PHANOTATE," scientists are predicting the protein structures and functions of bacteriophage genes so they can help researchers develop targeted therapeutics to treat bacterial infections.

The World's Most Abundant Predators

Sometimes referred to as microbial viruses, bacteriophages, literally meaning "bacteria eaters," are more abundant than any other living organism on Earth. Typical microbial viruses are composed of a protein capsid (DNA-filled head), an elgongated body with a collar, and a tail. These viruses have a penchant for a specific victim: bacteria. Unlike antibiotics, bacteriophages target specific bacteria species, and they evolve over time to keep up with their bacterial counterparts' mutations. "Bacteriophages have recognition molecules on their tails that help them identify their preferred bacterial host. They attach to the host and inject their DNA into the bacterial cell, turning the cell into a virus parts factory," explains Brian Souza, group leader for Biosecurity and Bioforensics



A bacteriophage, also known as a microbial virus, (a) magnified by a transmission electron microscope, (b) in 2D cross-section, consists of a DNA-containing head and a tail made up of protein, and (c) in 3D.

at Lawrence Livermore. Near the end of their reproductive cycle, bacteriophages produce holin, a protein that creates holes in a bacterium's cell wall. The bacteriophages then produce another protein called endolysin, which "lyses" or breaks the cell apart to release the newly generated bacteriophages so they can go on to infect other bacterial cells.

Genetic Testing for Bacteriophages

One of the first, biggest obstacles researchers faced in setting the groundwork for effective bacteriophage therapy was matching the particular bacterium infecting a person's body with the bacteriophage that would neutralize it. This matchmaking required characterization of various bacteriophages, particularly identification and annotation of their genes. "Before we began this project, a gene finder specifically for bacteriophages didn't exist. We had to use bacterial gene finders, which don't account for phage-specific characteristics," explains Carol Zhou, a computational biologist working on the PHANOTATE algorithm. "Phage genes are shorter, they're usually transcribed unidirectionally, and they're more compact than bacterial genes," says Zhou.

Antimicrobial Algorithm S&TR January 2021

Bacterial Genome



Phage Genome



Compared to the bacterial genome, the phage genome is much more compact and is usually transcribed in one direction.

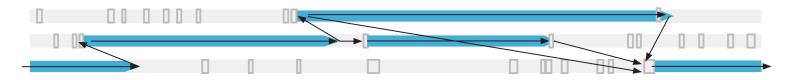
Phages also frequently exhibit unusual characteristics like gene overlap and nested genes within genes.

To develop a way to identify bacteriophage genes, the team created PHANOTATE, a gene caller machine learning algorithm specifically designed to identify phages. Given bacteriophages' diversity, unusual characteristics, and an abundance estimated to be 10³¹ (a number that translates into approximately one trillion phages for every grain of sand in the world), the computational burden necessary to curate massive datasets of phage genes made Lawrence Livermore particularly suited for the effort. "We decided our algorithm would model bacteriophage genomes with a weighted graph that has nodes and edges," says Zhou. "The nodes represent start and stop codons—sequences of three nucleotides which form a unit of genetic code—while the edges represent their translatable parts or open reading frames (ORFs), gaps, and strand switches. There's a weight penalty for anomalies, which the algorithm detects and works into the statistical likelihood of occurrence," explains Zhou.

The team implements the algorithm using the Bellman–Ford method, which treats the phage genome as a network of paths, with ORFs as the most favorable, and overlaps and gaps as less favorable. The network of connections portrayed by the weighted graph allows scientists to find the optimal gene path for picking the right bacteriophage to neutralize a particular bacterium. These paths are then compared to the results of established, bacteria-focused gene annotation tools to ensure PHANOTATE predicts accurate pathogen genomes. "Out of the four bacteria gene-calling codes we

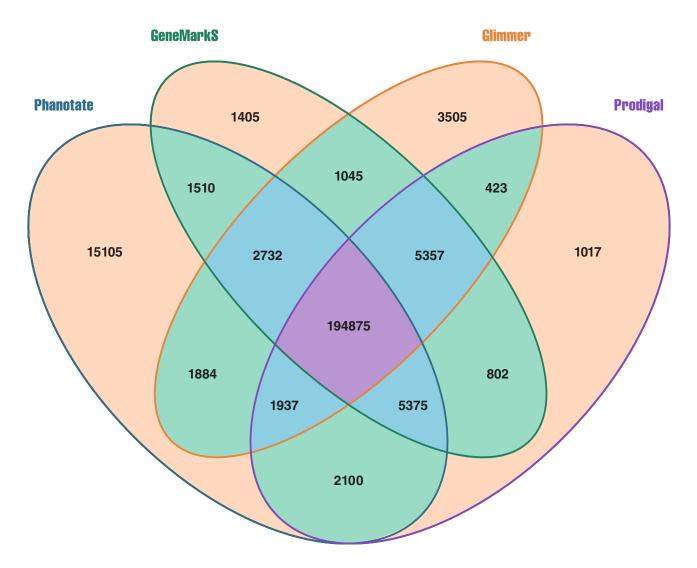
compare PHANOTATE against, our algorithm actually produces the largest total set of genes, including smaller genes that the other gene callers can't find," says Lawrence Livermore biomedical scientist Stephanie Malfatti. To ensure PHANOTATE does not produce false positives, the team leverages large databases like the National Center for Biotechnology Information (NCBI) Sequence Read Archive. "When we analyzed PHANOTATE's smaller genes, we found a number of good sequence matches and determined that not only can PHANOTATE find smaller genes than other gene callers, but it's also identifying them accurately," says Malfatti.

To locate and identify or "annotate" genes, the Livermore team also developed an automated throughput pipeline: the multiple-genome Phage Annotation Toolkit and Evaluator, or "multiPhATE." multiPhATE provides a scalable pipeline into which phage genomes can be entered, so their protein structures and predicted functions can be accessed by the biomedical research community. "The PHANOTATE algorithm results are put into the multiPhATE system and processed by annotation tools against a number of databases, several of which are specific to viruses," says Zhou. The combination of these two programs helps researchers identify the genes that, when translated into proteins, play a role in the infection and destruction process of a host cell. The team shares their codes on the open-source software platform GitHub, where the public has already identified potential enhancements. "One of the benefits of this platform is that other people can use and improve it," says Zhou.



A Bellman–Ford algorithm finds the optimal gene path for a phage.

S&TR January 2021 Antimicrobial Algorithm



Number of genes predicted by the four primary bacterial gene prediction algorithms: GeneMarkS, Glimmer, Prodigal, and PHANOTATE and combinations thereof, with PHANOTATE identifying the largest total set of genes. Orange background: predicted by a single algorithm; green background: predicted by two algorithms; blue background: predicted by three algorithms; purple background: predicted by all four algorithms.

Implementing Bacteriophage Therapy

Phages can be either lytic or temperate, so the team is focused on identifying lytic phages because they will quickly corrode and kill bacteria. Malfatti explains, "Gene annotation gives us the ability to look at specific suites of proteins and identify the genes that will enhance a phage's effectiveness, as well as the phages that will be most effective against a particular bacterium after comparative analyses."

With renewed funding, Souza's team of collaborative, multidisciplinary experts is working on refining and further enhancing the PHANOTATE algorithm even further, so it can identify additional gene overlaps and nested genes. Ultimately, the team anticipates that bacteriophage therapy

will one day be used in a wide range of areas, including infectious diseases, skin grafting, tissue repair, and gut health. Says Zhou, "For us, success is defined in stages. Reaching deployment and use of the code by other scientists or publishing another paper about our findings are hallmarks of success."

—Lauren Casonhua

Key Words: bacteriophage therapy, Forensic Science Center, gene annotation, microbial viruses, multiPhATE, National Center for Biotechnology Information (NCBI), pathogen, PHANOTATE.

For further information contact Brian Souza (925) 423-4642 (souza21@llnl.gov).